Reprinted from Journal of Food Science

HOWARD R. MOSKOWITZ and JOEL L. SIDEL[1]
Food Acceptance Lab., U.S. Army Natick Lab., Natick, MA 01760

MAGNITUDE AND HEDONIC SCALES OF FOOD ACCEPTABILITY

SUMMARY—Nine experiments were conducted to compare magnitude (ratio) and hedonic 9-point (category) scales of food acceptance. Five experiments were conducted with foods varied along one or more quantitative dimensions, and the remaining four with foods representing different flavors of a single product type. Both evaluation procedures appeared to be equally sensitive to differences in food acceptability, but each procedure provided additional information as well. Magnitude scales quantified the ratios of food acceptability among different items, and the hedonic scale provided numerical and verbal categories of acceptance. Each scale may be used to complement the other in the measurement of food acceptance.

INTRODUCTION

Category scaling

During the past two decades the 9-point hedonic scale, developed by the Quartermaster Food and Container Institute (Peryam and Pilgrim, 1957) has been a major tool for measuring food acceptability. The 9-point scale exemplifies the well-known category scale that psychologists have long used to measure attitudes (Guilford, 1954). In its particular form used to measure food acceptance the scale comprises nine categories which range from "dislike extremely" to "neutral" to "like extremely." The hedonic scale is thus anchored at three major categories, although the remaining six categories are assigned representative adjectives for the convenience of the panelists. The mathematical properties that make the scale useful for measurement are based upon the assumption that the psychological intervals between neighboring categories are equal across the entire scale, so that the user can employ several statistical tests to analyze the subjective estimates. Procedures such as analysis of variance have been used extensively to study variability in food preference judgments, and the hedonic scale has proven itself to be a simple and effective measuring instrument.

As an example of the category scale, the hedonic scale may be subject to the same limitations that apply to all scales of this type (Stevens and Galanter, 1957). These limitations derive from the mathematical properties of interval scales and from the systematic biases of human judges. Mathematically, the category scale lacks a true zero, so that one cannot conclude anything about ratios of acceptability, and may only conclude that two items either differ in acceptability or do not. In addition, judgment biases pervade all category scales (Stevens, 1957). Stevens has shown in numerous studies that judges tend to avoid using the extreme categories at both ends of such scales so that in effect they center their judgments around the middle categories (Stevens and Galanter, 1957). Thus, the nine categories of the hedonic scale may be effectively truncated to seven, so that the scale loses its usefulness for measuring certain degrees of food acceptability. Finally, the intervals between neighboring categories may be unequal (Peryam and Pilgrim, 1957), so that an interval of one category separating two highly acceptable foods may represent a greater psychological difference than the same interval for foods that are disliked, or liked only moderately.

Ratio Scaling—Magnitude Estimation

"Magnitude estimation" or free number matching has also been used in sensory evaluation (Stevens, 1953; Moskowitz, 1968). This form of scaling was developed at Harvard University, in the Psychoacoustic Laboratory, to assess the relation between physical intensity and the perceived magnitude of brightness and loudness. In many studies, more than a score of sensory and perceptual continua have been scaled by magnitude estimation, with results suggesting that the procedure of free number matching produces reliable and meaningful ratio scales of sensory magnitude. Other continua besides numbers have been used as measuring units (e.g., loudness, Stevens, 1966; handgrip, Stevens, Mack and Stevens, 1960), further suggesting that the sensory scales obtained by these matching procedures obey ratio properties. The method of magnitude estimation may provide answers to the objections about category scales just discussed.

In a scaling experiment with magnitude estimation the panelist is instructed to assign numbers to stimuli in proportion to the perceived intensity of his sensations. Because the judge is unconstrained as to the range and size of his numbers, the magnitude scale is not truncated at the extremes of sensory magnitude, and the scale that emerges has the properties of a ratio scale of magnitude. For example, the size of the numbers that the panelist uses does not influence the scale—rather, only the ratios between his numbers convey information. The starting stimulus may be any member of the stimulus sequence, and except for some minor biases the form of the ratio scale will be relatively unchanged.

Panelists are instructed to maintain numerical proportions, thus allowing a ratio scale of sensory intensity to be constructed for the particular dimension being studied. The goal of the scaling procedure is to place stimuli on a scale so that one stimulus can be rated as a specified number of times stronger or weaker than a reference stimulus. In addition, the judgments produce numerical values for sensory magnitude that may be correlated with measures of physical intensity in order to derive subjective-objective correlations. This type of correlation has been studied for more than two dozen perceptual continua (Stevens, 1960), and in great detail for the taste intensity of numerous taste stimuli and taste mixtures (Moskowitz, 1968; 1970a; 1970b).

The present study compares magnitude and hedonic scales for foods. Its purpose is to determine whether the procedure of magnitude estimation can be used as a technique for evaluating food acceptance, either in conjunction with, or independent of, the standard hedonic scale. Two basic comparisons between the procedures are: a) sensitivity in measuring differences in product acceptability, and b) the relation between acceptability and an underlying quantitative dimension of the food.

EXPERIMENTAL

TWO SERIES of studies were conducted. The first concerned foods that were prepared in the Acceptance Laboratory of the U.S. Army Natick Laboratories, where one or two dimensions of the food were systematically varied from the standard recipe. The second concerned several different varieties of the same food (e.g., different flavors of a beverage), prepared according to the manufacturer's instructions, without any alterations in the standard recipe. In both series, over-all acceptability of the samples was evaluated, but in the first series the functional relation was also determined between the acceptability of a food and an underlying quantitative dimension. The materials tested in the nine experiments are listed in Table 1, along with summary statistics about mean judgments and measures of variability.

In each experiment two sets of panelists were used; one group judged acceptability according to the standard hedonic scale; the

[1] Present address: Hunt-Wesson Foods, Fullerton CA 92634.

TABLE 1

Food	No. in each group	Variables	Magnitude scale: Geometric mean	Magnitude scale: Standard deviation*	Arithmetic mean	Hedonic scale Standard deviation: Prenormalization	Hedonic scale Standard deviation: Post-normalization
1. Karo syrup	24	Percentage syrup					
		80	0.85	0.20	6.42	1.38	0.69
		60	0.93	0.22	6.50	1.35	0.45
		40	1.23	0.28	6.17	1.34	0.64
No significant** ratio or difference							
2. Cherry drink	25	Sugar/beverage Ratio (1=Normal)					
		1.00/1.00	1.26	0.31	6.36	1.25	1.00
		1.00/0.66	1.51	0.36	7.08	0.86	1.07
		1.00/1.50	0.48	0.31	5.68	1.57	0.92
		0.66/1.00	0.56	0.50	5.52	1.61	1.10
		1.50/1.00	2.04	0.34	6.76	1.09	0.87
Least significant** ratio			2.0		Difference	0.9	0.7
3. Tuna spread	25	Tuna/Mayo Ratio (1=Normal)					
		2.00/1.00	0.51	0.34	6.04	1.70	1.21
		0.50/1.00	1.78	0.26	6.72	1.57	0.88
		1.00/1.00	0.96	0.17	5.92	1.85	0.80
		1.00/2.00	1.87	0.28	7.00	1.26	0.68
		1.00/0.50	0.58	0.31	5.52	2.00	0.96
Least significant** ratio			1.15			0.8	0.8
Recipe (a) 15 3/4 oz tuna fish (b) 13 1/2 tb mayonnaise (c) 1 1/2 tsp lemon juice (d) 5 shakes onion powder							
4. Chocolate chip cookies	25	Sugar/chocolate Ratio (1=Normal)					
		1.00/1.00	2.00	0.30	7.00	1.40	0.90
		0.25/1.00	1.00	0.10	6.70	1.50	0.90
		4.00/1.00	0.50	0.50	5.60	1.80	1.40
		1.00/0.33	0.79	0.30	6.40	1.70	1.00
		1.00/3.00	1.26	0.30	7.00	1.90	1.60
Least significant** ratio			2.0		Difference	1.1	1.1
5. Vegetable soup	25	Percent Soup					
		2.6	0.40	0.40	3.60	2.00	1.60
		5.2	0.79	0.50	6.20	2.00	1.60
		7.0	2.50	0.30	6.70	2.00	1.50
		10.5	2.50	0.40	6.90	1.80	1.50
		21.0	0.50	0.50	4.90	2.30	2.10
Least significant** ratio			2.5		Difference	1.5	1.5
6. Cookies	20	Flavor					
		Gaucho	1.10	0.30	7.80	1.00	0.60
		Fudge Town	0.90	0.10	7.00	1.00	0.50
		Lemon Punch	1.00	0.20	7.50	0.90	0.50
		Lickety Split	1.00	0.20	7.60	0.70	0.60
Least significant** ratio			None		Difference	0.5	0.5
7. Beverages	24	Flavor					
		Grape	1.10	0.30	6.7	1.20	0.60
		Lemon	1.00	0.30	6.9	1.30	0.80
		Orange	0.90	0.20	6.3	1.10	0.90
		Cherry	1.00	0.30	6.8	1.60	1.00
No significant** ratio or difference							
8. Cheese spreads		Flavor					
		Blue Cheese	1.00	0.30	6.80	1.50	1.50
		Whipped Cream Cheese	1.00	0.30	6.80	1.30	1.00
		Clam	0.80	0.30	6.30	1.90	1.30
		Onion	1.12	0.30	6.80	1.70	1.50
No significant** ratio or difference							
9. Jellies	20	Flavor					
		Grape jam	0.80	0.20	7.10	1.70	1.00
		Strawberry jam	0.70	0.30	7.30	1.40	0.70
		Grape jelly	1.20	0.20	7.70	0.90	0.80
		Strawberry jelly	1.60	0.30	8.30	0.70	0.70
Least significant** ratio			1.6		Difference	0.8	0.8

*After normalization.
**$P<0.05$.

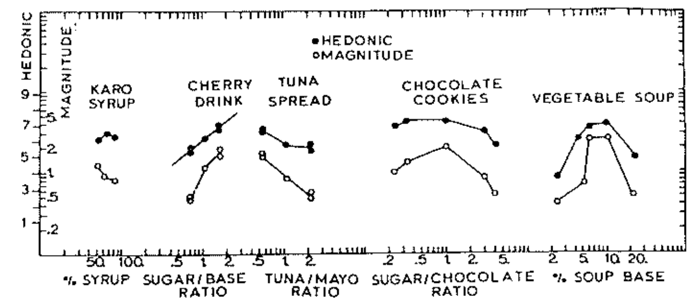

Fig. 1—Acceptability of foods whose ingredients were systematically varied in the laboratory. Open circles represent the geometric means of magnitude estimates, and are plotted in log-log coordinates. Closed circles represent the arithmetic means of hedonic judgments, and are plotted in semilog coordinates.

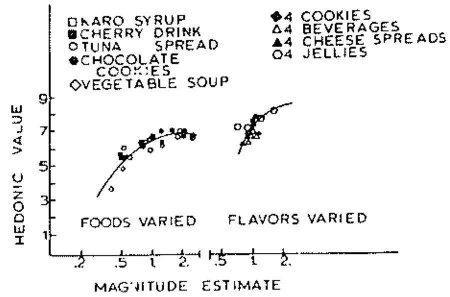

Fig. 2—Relation between the logarithm of magnitude estimation of acceptability (abscissa) and the hedonic judgment of acceptability. Each point represents one sample, and the values from magnitude estimation and 9-point hedonic scaling are mean values averaged across panelists.

other used magnitude estimation. Panelists in the second group were instructed to assign numbers to the samples in proportion to acceptability. The first number could be any one of the panelist's choice, except zero or negative. The remaining numbers were to be proportional to acceptability, with the standard being the first sample. Panelists were trained in magnitude estimation by being shown a series of lines and areas on index cards. They were instructed to assign numbers in proportion to apparent length and area, respectively, a simple task that illustrated the procedure of ratio scaling.

Usually 20–25 individuals served as panelists in each group of judges. The two groups were run in a single session (about 1.5 hr) to ensure equivalent food samples across groups. The samples were served successively and in small portions (2–3 oz). An interval of 30 sec elapsed between completion of one sample and presentation of the next, and the stimulus order was randomized for each judge in each experiment.

Analysis

Two analyses were run on the judgments. The first was an analysis of variance of the judgments to determine whether there was a significant treatment effect across samples. This was done for each of the nine experiments. The second analysis was run on those foods that were varied in the laboratory along one or more dimensions, and consisted of plotting the relation between acceptability scores and concentration.

Because the panelists in magnitude estimation were unconstrained in their choice of modulus (scale unit), the size of their numbers contributed a large proportion of the variance of the judgments. Variations in modulus may mask actual variability due to differences in acceptability, and can be reduced without affecting information about ratios of acceptability or differences in categories (for the hedonic scale). The magnitude estimates were first normalized by the procedure of multiplying the judgments of a single panelist for each sample in each session by a constant that made his geometric mean estimate across samples equal to 1.0. The ratios between estimates remained unchanged; therefore, the information about ratios of acceptability was preserved, but the size of the numbers of each panelist was brought into congruity with a standard magnitude level. In the same way the hedonic judgments were normalized by adding a constant to the judgments of each panelist so that the average judgment equalled zero. Since the same constant was added to all of the judgments of the panelist, the difference between categories is preserved. Again, all panelists were placed on the same part of the scale, and the variability of the judgments was reduced.

The analyses of variance were done on the hedonic ratings both before and after normalization, to determine whether the reduction in variability would affect final conclusions regarding differences in product acceptability. On the other hand, only one analysis of variance was performed on the results from the magnitude estimates. The logarithms of the estimates, rather than the estimates themselves, were subjected to analysis after normalization. The logarithms are the more appropriate measure because when subjective ratios of acceptability are converted to logarithms they become differences in acceptability. The additive models assumed in analysis of variance then become appropriate for these differences.

RESULTS & DISCUSSION

Results of analysis of variance for the entire nine experiments are presented in Table 1. In general, both methods of scaling for over-all acceptability concur in determining whether two samples differ significantly in acceptance rating; thus, the methods would be interchangeable if the only decision to be made were "accept-reject." In each experiment a difference was considered significant if the probability of that difference was less than 0.05.

The relation between physical concentration and the two scales is shown in Figure 1, in which the magnitude estimates are plotted in log-log coordinates, whereas the hedonic judgments themselves are plotted against log concentration. The two sets of functions suggest that concentration influences the acceptability of these food products, and that the two scales measuring acceptability resemble each other. Of importance, however, is the unique ability of the magnitude scale to measure how much more acceptable one food appears to be than another. The hedonic scale, in contradistinction, assesses only the category of acceptability, along an interval scale, so that the ratios of subjective acceptability cannot be obtained. Thus, the magnitude scale appears to yield subjective measurement more along the lines of traditional measurement of magnitude in physical and biological science.

Figure 2 presents the relation between the logarithm of the magnitude estimate and the hedonic judgment, across the entire set of samples investigated in the two sets of experiments. The hedonic judgments are the raw values obtained before normalization, whereas the logarithms of the magnitude estimates represent the normalized estimates. The relation is curvilinear over the entire range, but it steepens at the low end of acceptability (around a hedonic value of 5.0). Such a steepening might be predicted if one assumes as Stevens does (1957) that category scales have steeper slopes at low intensities. The neutral category of 5.0 is a low intensity of acceptance or rejection. As such, the neutral category is common to both ends of the hedonic scale, and is the origin from which both acceptance and rejection categories emerge. As a first approximation the relation between the category scale (C) and the magnitude scale (M) may be written in the logarithmic form

$$C = k \log (M + a)$$

in which the additive constant a corrects for the steepening at low magnitudes of acceptability. Eisler (1962) has presented an interesting discussion of the mathematical relation between the two scales for simpler perceptual continua.

The importance of the foregoing results is that the method of magnitude estimation with its ratio properties may be able to supplement category scaling in determining the degree of food acceptability. If both procedures are to be useful, then category scaling may be valuable in labelling the various perceptual categories of acceptance, whereas the magnitude scale may be employed in refining the measurement of the degree of acceptability. When an independent attribute of the food is systematically varied, the magnitude scale can become increasingly important in specifying the relation between over-all acceptability and physical concentration. This subjective-objective correlation using ratio scaling is far more meaningful than a similar correlation in which only a set of categories is used for the psychological scale.

REFERENCES

Eisler, H. 1962. Empirical test of a model relating magnitude and category scales. Scand. J. Psychol. 3: 88.

Guilford, J.P. 1954. "Psychometric Methods," Chap. 11. McGraw-Hill, New York.

Moskowitz, H.R. 1968. Scales of intensity for single and compound tastes. Ph.D. thesis, Harvard University, Cambridge, Massachusetts.

Moskowitz, H.R. 1970a. Ratio scales of sugar sweetness. Percept. & Psychophys. 7: 315.

Moskowitz, H.R. 1970b. Sweetness and intensity of artificial sweeteners. Percept. & Psychophys. 8: 40.

Peryam, D.R. and Pilgrim, F.J. 1957. Hedonic scale method of measuring food preferences. Food Technol. 9: 14.

Stevens, J.C., Mack, J.D. and Stevens, S.S. 1960. Growth of sensation on seven continua as measured by force of handgrip. J. Exptl. Psychol. 59: 60.

Stevens, S.S. 1953. On the brightness of lights and the loudness of sounds. Science 118: 576.

Stevens, S.S. 1957. On the psychophysical law. Psychol. Rev. 64: 153.

Stevens, S.S. 1960. The psychophysics of sensory function. Am. Scientist 48: 226.

Stevens, S.S. 1966. [illegible]ing functions between loudness and ten other continua. Percept. & Psychophys. 1: 5.

Stevens, S.S. and Galanter, E. 1957. Ratio scales and category scales for a dozen perceptual continua. J. Exptl. Psychol. 54: 377.

Ms received 10/7/70; revised 12/24/70; accepted 12/28/70.